D.P. van Velzen

Erdős-Rényi Random Graphs

Bachelor thesis

Thesis advisors:
prof.dr. W.Th.F den Hollander, dr. O. Jovanovski

July 4, 2017

Mathematisch Instituut, Universiteit Leiden

Contents

1 Introduction 2

2 Phase change behaviour 3
- 2.1 Comparison to branching processes . . . 4
- 2.2 Poisson branching processes . . . 6
- 2.3 The subcritical regime . . . 7
- 2.4 The supercritical regime . . . 10

3 Marginal laws for event counts 11
- 3.1 Relative entropy . . . 12
- 3.2 Conditioning on edge count . . . 12
- 3.3 Conditioning on degree sequence . . . 15

4 Conclusion 20

References 21

1 Introduction

Networks such as social media, world wide web and systems of transportation are ubiquitous in everyday life. The mathematical properties of such real world networks are extensively studied, and progress is rapid due to the ease of data analysis of networks. These networks often share interesting properties such as power law degree distributions and small world connectedness. From this perspective it is appropriate to look at real-world networks as if they were graphs.

Definition 1 *(Graph).* A simple undirected graph is an ordered pair $G = (V, E)$ of a set V of vertices and a set $E \subseteq V^2$ of edges, where an element $e \in E$ denotes a connection between two vertices. The elements of E are non-ordered pairs of vertices. «

Vertices may denote profiles in a social media context or servers in a www. context. Graphs that model such networks have comparable properties to theoretical graphs. These networks were not engineered to be similar, but have properties that naturally emerge. In many applications it is useful to look at the vertices and edges of graphs as if their location and connection were ruled by chance. One specific model of a random graph is the Erdős-Rényi model.

Definition 2 *(Erdős-Rényi).* Let $K_n = (V, E')$ be the complete graph on n vertices. We construct a random set $E \subset E'$ in the following manner. Let $p \in [0, 1]$, and for each element $e \in E'$, take $e \in E$ with probability p and $e \notin E$ with probability $1 - p$. The graph $ER_n(p) = (V, E)$ obtained in this way is called an Erdős-Rényi random graph with vertices $V = [n] = \{1, 2, \ldots, n\}$ and edges E. For most of section 2 we will look at graphs where $p = \lambda/n$. The law for these graphs will be denoted by $\mathbb{P}_\lambda$, with $\lambda \in \mathbb{R}^+$. [6] «

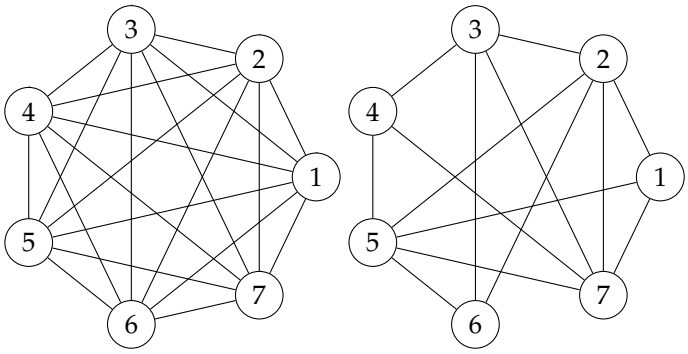

Figure 1: Erdős-Rényi graph for $p = \frac{1}{2}$, $n = 7$. Left: K_n. Right: $\mathrm{ER}_n(p)$

This specific model arises as a very simple example of a random graph. It does not have all properties that real-world networks usually have, but it is interesting for other reasons. The model has phase transitions in the limit $n \to \infty$ depending on the value of λ, which will be discussed in the next section.

2 Phase change behaviour

In this section we will give an outline of two proofs concerning connectedness for graphs that are subcritical ($\lambda < 1$), given in section 2.3, and supercritical ($\lambda > 1$), given in section 2.4. We will first give definitions that will be used in section 2.1 to compare the Erdős-Rényi random graph to branching processes, which will be defined in section 2.1. In section 2.2 we cite results needed for the proofs in sections 2.3 and 2.4. This section is based on chapter 4 from van der Hofstad's monograph [6] with verbatim definitions.

Notation: call $a, b \in V$ connected when $ab \in E$. Write $a \leftrightarrow b$ when there is a path of connected vertices from a to b. This means that there are vertices $m_1, \ldots, m_k$ such that $am_1, m_1m_2, \ldots, m_kb \in E$. By convention $a \leftrightarrow a$.

Definition 3 *(Cluster).* Let v be in the vertex set V of $ER_n(p)$. Define the *cluster* $\mathcal{C}(v)$ of v by

$$\mathcal{C}(v) = \{x \in V : v \leftrightarrow x\}.$$

«

Let $|\mathcal{C}(v)|$ denote the number of vertices in $\mathcal{C}(v)$. The largest cluster is any cluster for which $|\mathcal{C}(v)|$ is maximal, so that

$$|\mathcal{C}_{\max}| = \max_{v \in V} |\mathcal{C}(v)|. \tag{1}$$

We will now define a procedure called an *exploration process*, that is used to find the cluster of a certain vertex v in any given graph $G = (V, E)$. During the course of the exploration, each vertex in the graph has one of three statuses: active, neutral or inactive. At $t = 0$ only vertex v is active, all others are

neutral. At time $t \geq 1$ we take the vertex w with smallest assigned number and run over the edges ww', where w' runs over all neutral vertices. If ww' is in the edge set E, then we let w' become active. After all neutral vertices are searched, we set w to be inactive. When the process terminates at time t, the number of inactive vertices is the cluster size $|\mathcal{C}(v)| = t$.

```
Exploration of Fig. 1
t=0   1 active        none inactive          2 3 4 5 6 7 neutral
t=1   2 5 7 active    1 inactive             3 4 6 neutral
t=2   3 5 6 7 active  1 2 inactive           4 neutral
t=3   4 5 6 7 active  1 2 3 inactive         none neutral
t=4   5 6 7 active    1 2 3 4 inactive       none neutral
t=5   6 7 active      1 2 3 4 5 inactive     none neutral
t=6   7 active        1 2 3 4 5 6 inactive   none neutral
t=7   none active     1 2 3 4 5 6 7 inactive none neutral
```

We can relate the exploration process to a so called branching process as will be discussed in section 2.1 by setting S_t to be the number of active vertices at time t. This means that $S_0 = 1$. Let X_t denote the number of vertices that become active at time t. Then

$$S_0 = 1, \quad S_t = S_{t-1} + X_t - 1, \tag{2}$$

i.e., the number of active vertices equals the number of active vertices at time $t-1$ plus the number of vertices that becomes active minus the one that becomes inactive. This is true for any graph we explore, however, the probability distribution of X_t is specific to the Erdős-Rényi random graph. We see that

$$X_t \sim \text{Bin}(n - (t-1) - S_{t-1}, p)$$

is dependent on t and S_{t-1}. This means that the X_i for $i \in \mathbb{Z}_{>0}$ are not i.i.d.. If the X_i were i.i.d., then the process described by eq. (2) would be called a *branching process*, as will be explained in definition 4. The properties of branching processes will prove useful.

Let T denote the first t for which $S_t = 0$,

$$T = \min\{t : S_t = 0\}.$$

2.1 Comparison to branching processes

We now take another look at eq. (2) and compare the exploration of a random graph to the exploration of a stochastic process called *branching process* where $X_i \sim \text{Bin}(n, p)$ i.i.d. for all $i \in \mathbb{Z}_{>0}$.

Definition 4 *(branching process)***.** A branching process is a stochastic process that gives rise to a layered graph where each layer represents a generation of individuals. Each individual in a generation has a stochastic number of offspring, giving rise to the next generation. In mathematical terms: Let Z_n be the number of individuals in the n-th generation, and $(\zeta_{n,i})_{n\geq 0, i>0}$ a two-dimensional array of i.i.d. stochastic variables representing the offspring of

the nodes in each generation. The recursive relation

$$Z_0 = 1 \quad Z_n = \sum_{i=1}^{Z_{n-1}} \zeta_{n,i}$$

defines a branching process. «

We visualize a branching process as a tree (graph) where each Z_n denotes the number of vertices that are at distance n from the root. Exploring such a graph with the exploration process and starting in the root, we get

$$S_0 = 1, \quad S_t = S_{t-1} + X_t - 1,$$

but for the exploration of a branching process the X_t would be i.i.d.

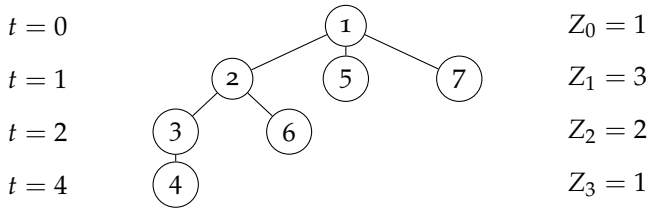

Figure 2: This tree has the exact same exploration as the graph in Fig.1.

We will establish stochastic bounds on the exploration of an Erdős-Rényi random graph in terms of branching processes. These bounds will allow us to use powerful results for branching processes.

Theorem 5 *(Stochastic domination of the cluster size).* *For every $k \geq 1$ and $v \in [n]$*

$$\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) \leq \mathbb{P}_{n,p}(T^{\geq} \geq k), \quad \textit{i.e.,} \quad |\mathcal{C}(v)| \preceq T^{\geq},$$

where $T^{\geq}$ is the total progeny of a binomial branching process as described in section 2.1 with parameters n and p. «

Proof. Let $n_i = n - i - S_i$ be the number of neutral vertices at $t = i$ in the exploration. Then the exploration as described in eq. (2) has $X_i \sim \text{Bin}(n_{i-1}, p)$. Let $Y_i \sim \text{Bin}(n - n_{i-1}, p)$ and define

$$X_i^{\geq} := X_i + Y_i.$$

By Vandermonde's identity we see that $X_i^{\geq} \sim \text{Bin}(n, p)$ is independent of n_{i-1} for all i. Also $X_i \leq X_i^{\geq}$ because $Y_i \geq 0$. Now let

$$S_i^{\geq} := X_1^{\geq} + \ldots + X_i^{\geq} - (i-1).$$

Then

$$\begin{aligned}\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) &= \mathbb{P}(\forall t \leq k-1 : S_t > 0) \leq \mathbb{P}(\forall t \leq k-1 : S_t^{\geq} > 0) \\ &= \mathbb{P}(T^{\geq} \geq k),\end{aligned}$$

where $T^{\geq} := \min\{t : S_t^{\geq} = 0\}$ is the total progeny for a binomial branching process with parameters n, p. □

Theorem 6 *(Lower bound on cluster tail).* *For every $k \in [n]$,*

$$\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) \geq \mathbb{P}_{n-k,p}(T^{\leq} \geq k),$$

where $T^{\leq}$ is the total progeny of a binomial branching process with parameters $n-k$ and p. «

Note: Since $n-k$ on the R.H.S. depends on k, Theorem 6 does not imply a stochastic lower bound on $|\mathcal{C}(v)|$.

Proof. Using the same definition for n_i as in the proof of Theorem 5, we define $\mathcal{T}_k$ to be the stopping time

$$\mathcal{T}_k := \min\{t : n_t \leq n-k\},$$

denoting the first time t when a cluster of at least size k has been found. Therefore $\mathcal{T}_k \leq k-1$, since at time $t = k$ the number of inactive vertices is k, so at time $t = k-1$ we either have no active vertices and $\mathcal{T}_k$ is not defined, or we have a number of active vertices and $\mathcal{T}_k \leq k-1$. We can trivially rewrite

$$\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) = \mathbb{P}_{n,p}(\forall t \leq \mathcal{T}_k : S_t > 0).$$

As in the proof of Theorem 5, we use Vandermonde's identity to relate the exploration process with $X_i \sim \text{Bin}(n_{i-1}, p)$ for all i to a branching process with i.i.d. sequence $X_i^{\leq} \sim \text{Bin}(n-k, p)$. We do this by defining $Y_i \sim \text{Bin}(n_{i-1} - (n-k), p)$ and noting that

$$X_i = X_i^{\leq} + Y_i, \quad \forall i \leq \mathcal{T}_k.$$

Since $Y_i \geq 0$ we see that $X_i \geq X_i^{\leq}$. Define

$$S_i^{\leq} = X_1^{\leq} + \ldots + X_i^{\leq} - (i-1).$$

Then $S_t^{\leq} \leq S_t$ for all $t \leq \mathcal{T}_k$. We can now bound

$$\begin{aligned}
\mathbb{P}_{n,p}(|\mathcal{C}(v)| \geq k) &= \mathbb{P}_{n,p}(\forall t \leq \mathcal{T}_k : S_t \geq 0) \\
&\geq \mathbb{P}_{n,p}(\forall t \leq \mathcal{T}_k; S_t^{\leq} > 0) \\
&\geq \mathbb{P}_{n,p}(\forall t \leq k-1 : S_t^{\leq} > 0) \\
&= \mathbb{P}_{n-k,p}(T^{\leq} \geq k).
\end{aligned}$$

□

Theorems 5 and 6 give us upper and lower bounds on the exploratory process in terms of branching processes.

2.2 Poisson branching processes

For a branching process where $X_{n,i}$ are i.i.d. Poisson distributed with parameter λ we write T^* for the total progeny, i.e., $T^* = \sum_{n=0}^{\infty} Z_n$. We use $\mathbb{P}_\lambda^*$ to

denote the law of this branching process. We state the following results from chapter 3.6 of [6]:

$$\mathbb{P}^*(k \leq T^* < \infty) \leq e^{-I_\lambda k}, \tag{3}$$
$$I_\lambda = 1 - \lambda - \log(\lambda).$$

Theorem 7 *(Poisson and binomial branching processes).* *For the branching process with binomial offspring distribution with parameters n and p, and the branching process with Poisson offspring distribution with parameter $\lambda = np$, for each $k \geq 1$,*

$$\mathbb{P}_{n,p}(T \geq k) = \mathbb{P}^*_\lambda(T^* \geq k) + c_n(k),$$

where T and T^* are the total progenies of the binomial and Poisson branching processes, respectively, and

$$|c_n(k)| \leq \frac{\lambda^2}{n} \sum_{s=1}^{k-1} \mathbb{P}^*_\lambda(T^* \geq s).$$

In particular, $|c_n(k)| \leq k\lambda^2/n$. «

2.3 The subcritical regime

We will now establish bounds for the size of the largest cluster in the subcritical regime, i.e., when $\lambda = np < 1$. We will use some of the theory summarized in section 2.2.

Theorem 8 *(Upper bound on the largest subcritical component).* *Fix $\lambda < 1$. Then for every $a > 1/I_\lambda$, there exists a $\delta = \delta(a, \lambda) > 0$, such that*

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| \geq a \log(n)) = O(n^{-\delta}).$$

«

Proof. By Theorem 5 we know that for fixed $v \in V$,

$$P_{n,p}(|\mathcal{C}(v)| \geq k) \leq \mathbb{P}_{n,p}(T^{\geq} \geq k),$$

where $T^{\geq}$ is the total progeny of the binomial branching process defined by $n, p = \lambda/n$. As in the proof of Theorem 5, we define

$$S_t^{\geq} := X_1^{\geq} + \ldots + X_t^{\geq} - (t-1).$$

By Theorem 5 we have

$$P_{n,p}(|\mathcal{C}(v)| \geq k) \leq \mathbb{P}_{n,p}(S_t^{\geq} > 0) = \mathbb{P}_{n,p}(X_1^{\geq} + \ldots + X_t^{\geq} > t).$$

Using the Poisson limit theorem, we can bound this by

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k) \leq \mathbb{P}_{n,p}(X_1^{\geq} + \ldots + X_t^{\geq} > t) \leq e^{-tI_\lambda}.$$

By using the union bound $\mathbb{P}(\cup_i A_i) \leq \sum_i \mathbb{P}(A_i)$, we know that the probability that all clusters are larger than $t = a\log(n)$ is less than or equal to the sum of the probabilities that the clusters of every vertex are larger than t. Therefore

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| > a\log(n)) \leq n\mathbb{P}_\lambda(|\mathcal{C}(v)| > a\log(n)) \leq ne^{a\log(n)I_\lambda} = n^{1-aI_\lambda} = n^{-\delta}$$

with $\delta = aI_\lambda - 1 > 0$ for $a > 1/I_\lambda$. □

Theorem 9 *(Lower bound on the largest subcritical component).* *Fix $\lambda < 1$. Then for every $a < 1/I_\lambda$, there exists a $\delta = \delta(a,\lambda) > 0$, such that*

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| \leq a\log(n)) = O(n^{-\delta}).$$

«

Before starting the proof of Theorem 9 we will have to define some new stochastic variables and prove a proposition. Denote the number of vertices contained in clusters of size at least k by

$$Z_{\geq k} = \sum_{i\in[n]} \mathbb{1}_{\{|\mathcal{C}(i)|\geq k\}}.$$

We can relate $Z_{\geq k}$ to $|\mathcal{C}_{\max}|$ by noting that whenever $Z_{\geq k}$ is larger than k, at least k vertices are contained in a cluster of size at least k and therefore $|C_{\max}| \geq k$. Conversely, when $|C_{\max}| \geq k$ there is at least one cluster of size at least k and therefore there are at least k vertices in a cluster of size k. We write

$$\{|C_{\max}| \geq k\} = \{Z_{\geq k} \geq k\}.$$

Let

$$\chi_{\geq k}(\lambda) = \mathbb{E}_\lambda\left[|\mathcal{C}(v)|\mathbb{1}_{\{|\mathcal{C}(v)|\geq k\}}\right].$$

Proposition 10 *(Variance estimate for $Z_{\geq k}$).* *For every n and $k \in [n]$ and $\lambda > 0$,*

$$Var(Z_{\geq k}) \leq n\chi_{\geq k}(\lambda).$$

«

Proof. Using the definition of variance we see that

$$\mathrm{Var}(Z_{\geq k}) = \sum_{i,j\in[n]} \left[\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k) - \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k)^2\right].$$

To examine $\mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, |\mathcal{C}(j)| \geq k)$ we distinguish two cases: either $i \leftrightarrow j$ or not. Looking at the case where i and j are not connected, we make the following observation. W.L.O.G. if $|\mathcal{C}(i)| = l$, then

$$\mathbb{P}_{n,\lambda}(|C(j)| \geq k, i \not\leftrightarrow j) = \mathbb{P}_{n-l,\lambda}(|\mathcal{C}(j)| \geq k).$$

Since the probability of finding a large cluster increases with n, we see that for every $l \geq 1$,

$$\mathbb{P}_{n,\lambda}(|\mathcal{C}(i)| = l, |\mathcal{C}(j)| \geq k, i \not\leftrightarrow j) - \mathbb{P}_{n,\lambda}(|\mathcal{C}(i)| = l)\mathbb{P}_{n,\lambda}(|\mathcal{C}(j)| \geq k) \leq 0.$$

This in turn means that

$$\mathrm{Var}(Z_{\geq k}) \leq \sum_{i,j\in[n]} \mathbb{P}_\lambda(|\mathcal{C}(i)| \geq k, i \leftrightarrow j).$$

Hence

$$\begin{aligned}
\mathrm{Var}(Z_{\geq k}) &\leq \sum_{i\in[n]} \sum_{j\in[n]} \mathbb{E}_\lambda[\mathbb{1}_{\{|\mathcal{C}(i)|\geq k\}} \mathbb{1}_{\{j\in\mathcal{C}(i)\}}] \\
&= \sum_{i\in[n]} \mathbb{E}_\lambda[|\mathcal{C}(i)|\,\mathbb{1}_{\{|\mathcal{C}(i)|\geq k\}}] \\
&= n\mathbb{E}_\lambda[|\mathcal{C}(1)|\,\mathbb{1}_{\{|\mathcal{C}(1)|\geq k\}}] \\
&= n\chi_{\geq k}(\lambda).
\end{aligned}$$

□

We can now give the proof of Theorem 9.

Proof. In terms of $Z_{\geq k}$ we can restate Theorem 9 as $\mathbb{P}(Z_{\geq k_n} = 0) = O(n^{-\delta})$ with $k_n = a\log(n)$ and $a \leq 1/I_\lambda$.

The proof uses the Paley-Zigmund inequality: When X is integer valued and $\mathbb{E}(X) \geq m$, then

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{m^2}.$$

We will use the inequality for $Z_{\geq k_n}$. This means we must establish a lower bound for $\mathbb{E}(Z_{\geq k})$ and an upper bound for $\mathrm{Var}(Z_{\geq k})$. We use that for any $v \in V$

$$\mathbb{E}_\lambda[Z_{\geq k}] = nP_{\geq k}(\lambda), \quad \text{where} \quad P_{\geq k}(\lambda) = \mathbb{P}(|\mathcal{C}(v)| \geq k).$$

Take $k_n = a\log(n)$. We use the lower bound on the cluster size as established in Theorem 6. Again T is the total progeny of the binomial branching process with parameters $n - k_n, p$:

$$P_{\geq k_n}(\lambda) \geq \mathbb{P}_{n-k_n,p}(T \geq k_n).$$

By Theorem 7 we can approximate the binomial distribution by a Poisson distribution with mean $\lambda_n = \lambda(n - k_n)/n$,

$$\mathbb{P}_{n-k_n,p}(T \geq a\log(n)) = \mathbb{P}^*_{\lambda_n}(T^* \geq a\log(n)) + O\left(\frac{a\lambda^2\log(n)}{n}\right).$$

Using Stirling's approximation we can simplify the Poisson distribution to show that with $k_n = a\log n$ and any α such that $0 < \alpha \leq 1 - I_\lambda a$,

$$\mathbb{E}_\lambda[Z_{\geq k_n}] \geq n^\alpha.$$

Using Proposition 10 and analyzing $\chi_{\geq k_n}$, we get

$$\mathrm{Var}(Z_{\geq k_n}) \leq n\chi_{\geq k_n}(\lambda) \leq O(k_n n^{1-\alpha I_\lambda}).$$

We now have the required bounds to use the Payley-Zygmund inequality and estimate

$$\mathbb{P}_\lambda(Z_{\geq k_n} = 0) \leq \frac{\mathrm{Var}(Z_{\geq k_n})}{\mathbb{E}_\lambda[Z_{\geq k_n}]^2} \leq O(k_n n^{1-aI_\lambda-2\alpha}) = O(n^{-\delta}).$$

for n large enough and any $\delta < 2\alpha - (1 - I_\lambda a)$, where $2\alpha(1 - I_\alpha a) > 0$ when $0 < \alpha \leq 1 - I_\lambda a$. Hence

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| \leq a\log(n)) = \mathbb{P}_\lambda(Z_{\geq k_n} = 0) = O(n^{-\delta}).$$

□

Thus we have established strong bounds on $|\mathcal{C}_{\max}|$ in the subcritical regime. In particular we find that $|\mathcal{C}_{\max}| \sim a\log(n)$ for large n.

2.4 The supercritical regime

For the supercritical regime ($\lambda > 1$) we will show that all vertices in clusters of sufficient size are connected in one *giant* cluster. The cluster sizes are closely related to Poisson branching processes. For $\lambda > 1$ a positive ratio of branching processes will survive, i.e., $\forall t : Z_t > 0$. This tells us that there are large clusters. In fact $|\mathcal{C}_{\max}|$ will have a size of $O(n)$. We will make these claims more precise by summarizing four propositions that lead to the following theorem. Write $\zeta_\lambda = 1 - \eta_\lambda$ for the survival probability of a Poisson branching process with mean offspring λ.

Theorem 11 *(Law of large numbers for the giant component).* *Fix $\lambda > 1$. Then, for every $\nu \in (\frac{1}{2}, 1)$, there exists a $\delta = \delta(\nu, \lambda) > 0$ such that*

$$\mathbb{P}_\lambda(|\mathcal{C}_{\max}| - \zeta_\lambda n| \geq n^\nu) = O(n^{-\delta}).$$

«

We will first show, with the help of the Poisson branching process, that there is a positive probability of finding clusters that are much larger in size than the clusters found in the subcritical regime.

Proposition 12 *(Cluster tail is branching process survival probability).* *Fix $\lambda > 1$ Then, for $k_n \geq a\log n$ with $a > 1/I_\lambda$ and $I_\lambda = 1 - \lambda - \log(\lambda)$ and for n sufficiently large,*

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) = \zeta_\lambda + O(k_n/n).$$

«

This proposition can be proven by bounding $\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n)$ with the help of the binomial branching process as in Theorems 5 and 6. These can in turn be compared to Poisson branching processes via Theorem 7. Note that for the lower bound we will have to compare, using $\lambda_n = \lambda(1 - k_n/n)$,

$$\mathbb{P}_\lambda(|\mathcal{C}(v)| \geq k_n) \geq \mathbb{P}_{n-k_n,\lambda/n}(T_{\geq} \geq k_n) \geq \mathbb{P}^*_{\lambda_n}(T^* \geq k_n) + O(k_n/n).$$

In other words, we bound by the probability of another such a cluster appearing in the remaining graph. This gives us the bound stated in Proposition 12.

We will next use a method similar as for the variance bound in the subcritical regime. Let

$$\chi_{<k}(\lambda) = \mathbb{E}[|\mathcal{C}(v)|\mathbb{1}_{\{|\mathcal{C}(v)|<k\}}].$$

Proposition 13 *(Variance estimate of $Z \geq k$).* *For every n and $k \in [n]$,*

$$Var_\lambda(Z_{\geq k}) \leq (\lambda k + 1)n\chi_{<k}(\lambda).$$

«

Again, by splitting the variance according to whether the vertices are connected or not, we can derive this bound. Using this bound, we can again use Paley-Zygmund's inequality to show that

$$\mathbb{P}_\lambda(Z_{\geq k_n} - \zeta_\lambda n| \geq n^\nu) = O(n^{-\delta}).$$

This means that a positive fraction of vertices is in clusters of size at least k_n with high probability.

In a more technical proof, a bound can be derived on the number of vertices in clusters of sizes that are between the subcritical $O(\log(n))$ and the supercritical $O(n)$ sizes. First a rate function is defined as

$$J(\alpha, \lambda) = I_{g(\alpha,\lambda)}, \quad \text{with} \quad g(\alpha, \lambda) = (1 - e^{-\lambda\alpha})/\alpha.$$

Using this rate function and section 2.2, we obtain the following proposition.

Proposition 14 *(Exponential bound for supercritical clusters smaller than $\zeta_\lambda n$).* *Fix $\lambda > 1$ and let k_n be such that $k_n \to \infty$ as $n \to \infty$. Then, for any $\alpha < \zeta_\lambda$,*

$$\mathbb{P}_\lambda(k_n \leq |\mathcal{C}(v)| \leq \alpha n) \leq e^{-k_n J(\alpha,\lambda)}/[1 - e^{-J(\alpha,\lambda)}].$$

«

This means that the region of cluster sizes between the subcritical and supercritical sizes is rather empty.

Finally, it can be shown that

$$\mathbb{P}_\lambda(||\mathcal{C}_{\max}| - \zeta_\lambda n| \leq n^{-\nu}) \geq 1 - O(n^{-\delta}),$$

which means that the number of vertices in a large cluster is equal to n times the survival probability of the binomial branching process.

3 Marginal laws for event counts

In this section we will consider probability distributions on sets of Erdős-Rényi random graphs that satisfy a constraint on the total number of edges or on the degree sequence. In particular, we will use the definition of relative entropy to compare the probability distribution of the constrained graphs to that of the unconstrained graphs. The ideas for this section are from [2, 5].

3.1 Relative entropy

To compare the marginal laws on edge counts we will be using the following pseudo-distance.

Definition 15 *(Relative entropy).* For discrete probability distributions P and Q, we define the relative entropy $h(P|Q)$ of P relative to Q as

$$h(P|Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right).$$

We say, by convention, that for $P(i) = 0$, $P(i)\log(P(i)/Q(i)) = 0$. «

Note that the relative entropy is a pseudo-distance since it is non-symmetric and does not satisfy the triangle inequality. The relative entropy is non-negative, and zero if and only if $P = Q$.

3.2 Conditioning on edge count

We will now make explicit what is meant by conditioning on edge count. First for $n \in \mathbb{Z}_{>0}$ we call $\mathcal{G}_n$ the set of all graphs with n vertices. For $p \in (0,1)$, let $\mathbb{P}_p$ be the law of Erdős-Rényi random graphs of size n with parameter p. We define the following distributions for p and $G \in \mathcal{G}_n$, by conditioning on the total number of edges $|E|$:

$$P(G) = \mathbb{P}_p(G||E| = L),$$
$$Q(G) = \mathbb{P}_p(G).$$

Where $L \in \mathbb{Z}_{\geq 0}$. This means that P is the law on Erdős-Rényi graphs with parameter p that have exactly $L = p\binom{n}{2}$ edges. Q is the law of Erdős-Rényi graphs with parameter p.

For ease of notation, we will use the shorthand and notation $M = \binom{n}{2}$. We note that the number of graphs in $\mathcal{G}_n$ with exactly L edges is $\binom{M}{L}$. For Erdős-Rényi random graphs the probability of an edge being in the edge set E is uniformly distributed. This means that

$$P(G) = \mathbb{P}_p(G : L = pM) = \binom{M}{L}^{-1} \mathbb{1}_{\{|E|=L\}}.$$

We now examine the relative entropy

$$h(P|Q) = \sum_{G\in\mathcal{G}_n} \binom{M}{L}^{-1} \mathbb{1}_{\{|E|=L\}} \log\left(\frac{\binom{M}{L}^{-1}\mathbb{1}_{\{|E|=L\}}}{Q(G)}\right).$$

Since $\mathbb{1}_{\{|E|=L\}} = 0$ when $|E| \neq L$, we have that

$$h(P|Q) = \sum_{\substack{G\in\mathcal{G}_n \\ |E|=L}} \binom{M}{L}^{-1} \log\left(\frac{\binom{M}{L}^{-1}}{Q(G)}\right).$$

Here we only sum over graphs $G \in \mathcal{G}_n$ with $L = pM$, which means that $Q(G)$ is the same for all G and $Q(G) = p^L(1-p)^{M-L}$. Since our summands are no longer dependent on G, we sum over $\binom{M}{L}$ graphs and find

$$\begin{aligned} h(P|Q) &= \log\left(\frac{\binom{M}{L}^{-1}}{p^L(1-p)^{M-L}}\right) \\ &= -\log\binom{M}{L} - L\log(p) - (M-L)\log(1-p). \end{aligned}$$

We want to know more about the behavior of $h(P|Q)$. Stirling's approximation says that

$$\text{For } n \to \infty, \quad n! = n^n e^{-n}\sqrt{2\pi n}(1+o(1)).$$

We first take a look at

$$\log\binom{M}{L} = \log\left(\frac{M!}{L!(M-L)!}\right).$$

Note that as $n \to \infty$ each of $M, L, M-L$ tends to infinity. Therefore

$$\log\binom{M}{L} = \log\left[\frac{M^M e^{-M}\sqrt{2\pi M}}{L^L e^{-L}\sqrt{2\pi L}(M-L)^{M-L}e^{-(M-L)}\sqrt{2\pi(M-L)}}\,(1+o(1))\right],$$

where we note that the exponentials cancel out. We write $M^M = \frac{1}{M^{-L}M^{-(M-L)}}$ and use this to write

$$\begin{aligned} \log\binom{M}{L} &= \log\left[\frac{1}{(\frac{L}{M})^L(1-\frac{L}{M})^{(M-L)}}\sqrt{\frac{M}{2\pi L(M-L)}}(1+o(1))\right] \\ &= \log\left[\frac{1}{p^L(1-p)^{(M-L)}}\sqrt{\frac{M}{2\pi L(M-L)}}(1+o(1))\right] \\ &= -L\log(p) - (M-L)\log(1-p) - \frac{1}{2}\log\left[\frac{M}{2\pi L(M-L)}\right] + o(1). \end{aligned}$$

Returning to relative entropy, we find

$$\begin{aligned} h(P|Q) &= \frac{1}{2}\log\left[\frac{M}{2\pi L(M-L)}\right] + o(1) \\ &= -\frac{1}{2}\log(2\pi p(1-p)M) + o(1). \end{aligned}$$

Note that, since $M = \binom{n}{2} = n(n-1)/2$, we find that $\lim_{n\to\infty} n^{-1}h(P|Q) = 0$. This means that, for n large enough, P and Q are indiscernible, when viewed from single nodes. We call P and Q *measure equivalent*.

What happens when we take

$$Q(G) = \mathbb{P}_{p'}(G),$$

for $p' \in (0,1)$ and p' and p not necessarily equal? We then find a different relative entropy, namely,

$$h(P|Q) = -\frac{1}{2}\log(2\pi p(1-p)M) + M\left[p\log(\frac{p}{p'}) + (1-p)\log(\frac{1-p}{1-p'})\right] + O(1).$$

Therefore we find that the second term only vanishes when $p = p'$ and so P and Q are asymptotically equivalent if and only if $p = p'$. We next want to know if these calculations are still valid for the sparse case where

$$p = \lambda/n,$$
$$p' = \lambda'/n.$$

Remember that, for Stirling's approximation we required that each of M, L, $M - L$ goes to infinity. Since $L = pM = \frac{\lambda}{n}\binom{n}{2}$, we see that we are indeed allowed to use Stirling's approximation. In fact we see that we can take Stirling's approximation of the factorial for any $p = \lambda/n^\alpha$ as long as $\alpha \in (0,2)$. We again look at the order of growth of the relative entropy. First consider the first term. For $n \to \infty$

$$-\frac{1}{2}\log\left(2\pi\frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)M\right) + O(1) = -\frac{1}{2}\log(2\pi\lambda n) + O(1)$$

The error made by setting $M = n^2$ is still $O(1)$. For the other terms we make use of the Taylor expansion. Writing

$$-\log(1-x) = \sum_{k=1}^{\infty}\frac{x^k}{k}, \quad |x| \leq 1,$$

we get

$$\begin{aligned}
&M\left[p\log\left(\frac{p}{p'}\right) + (1-p)\log\left(\frac{1-p}{1-p'}\right)\right] + O(1)\\
&= M\left[\frac{\lambda}{n}\log\left(\frac{\lambda}{\lambda'}\right) + \left(1-\frac{\lambda}{n}\right)\log\left(\frac{1-\frac{\lambda}{n}}{1-\frac{\lambda'}{n}}\right)\right] + O(1)\\
&= \frac{n^2-n}{2}\left[\frac{\lambda}{n}\log\left(\frac{\lambda}{\lambda'}\right) + \left(1-\frac{\lambda}{n}\right)\left[\frac{(\lambda'-\lambda)}{n} + \frac{(\lambda'^2-\lambda^2)}{2n^2}\right]\right] + O(1)\\
&= \frac{n}{2}\left[\lambda\log\left(\frac{\lambda}{\lambda'}\right) + \left(1-\frac{\lambda}{n}\right)\left[(\lambda'-\lambda) + \frac{(\lambda'^2-\lambda^2)}{2n}\right]\right] + O(1).
\end{aligned}$$

Hence

$$h(P|Q) = -\frac{1}{2}\log(2\pi\lambda n) + \frac{n}{2}\left[\lambda\log\left(\frac{\lambda}{\lambda'}\right) + \left(1-\frac{\lambda}{n}\right)\left[(\lambda'-\lambda) + \frac{(\lambda'^2-\lambda^2)}{2n}\right]\right] + O(1).$$

Again n can be replaced by n^α with $\alpha \in (0,2)$. This means that, for $\alpha < 1$, P and Q are measure equivalent in the sparse case. For $1 \leq \alpha < 2$, on the other hand, P and Q are only measure equivalent when $\lambda = \lambda'$.

3.3 Conditioning on degree sequence

Next we take a look at the distribution of *degrees* in a graph.

Definition 16. The *degree* $\deg(a)$ of the vertex a is defined as

$$\deg(a) = \#\{x \in V \backslash \{a\} : ax \in E\}.$$

After labeling the vertices $i \in V = [n]$, we can use the notation $k_i^* = \deg(i)$. This gives us the *degree sequence* $C(G) = (k_1^*, \ldots, k_n^*)$ for the graph G, where $C(G)$ is a vector that gives us an ordered list of the degrees. «

The degree sequence can be viewed as a vector valued function $\vec{C}$ on the space of graphs $\mathcal{G}_n$. We want to see if the so-called *microcanonical ensemble* is measure equivalent to the so-called *canonical ensemble* of graphs.

Definition 17 *(Microcanonical ensemble).* The *microcanonical ensemble* of graphs is a uniform distribution P_{mic} on $\mathcal{G}_n$ satisfying the hard constraint

$$\vec{C}(G) = \vec{k}^*.$$

«

Definition 18 *(Canonical ensemble).* The *canonical ensemble* of graphs is the distribution P_{can} on $\mathcal{G}_n$, given by

$$\mathbb{P}_{\text{can}}(G) = \frac{e^{-\vec{\theta}\cdot\vec{C}(G)}}{Z(\vec{\theta})},$$

where $Z(\vec{\theta})$ is a normalizing factor and $\vec{\theta}$ is to be chosen such that

$$\sum_{G \in \mathcal{G}_n} \vec{C}(G)\mathbb{P}_{\text{can}}(G) = \vec{k}^*.$$

«

The latter ensures that the expected value of the degree of vertex i with respect to $\mathbb{P}_{\text{can}}$ is equal to k_i^* for every $i \in [n]$.

The specific forms of the microcanonical and the canonical ensemble ensure that the Shannon entropy is maximal.

Definition 19 *(Shannon entropy).* For $\mathbb{P}$ a probability distribution on $\mathcal{G}_n$ the *Shannon entropy* $S(\mathbb{P})$ is defined as

$$S(\mathbb{P}) = -\sum_{G \in \mathcal{G}_n} \mathbb{P}(G)\log(\mathbb{P}(G)).$$

«

Shannon entropy is a quantity introduced by Shannon in [4]. It satisfies three properties that make it useful in terms of information theory. In a system that has a set of possible events with probability $(p_1, p_2, \ldots)$, the Shannon entropy $S(p_1, p_2, \ldots)$ is the only measure that satisfies

1. S is continuous in each of the p_i.
2. $S(1/n, 1/n, \ldots, 1/n)$ increases monotonically with n, i.e., the entropy of a roll with a dice is higher than flip with a coin.
3. Probabilities of events that can be described as consecutive choices are multiplicative (see [4] for more details).

The Shannon entropy is the only measure (modulo multiplication by constants) that satisfies these constraints, which makes it a very fundamental quantity.

In an article by Jaynes [3] it is argued that the values of $(p_1, p_2, \ldots)$ of events where we only know the value of macroscopic constraints, such as degree sequence, always maximize Shannon entropy. This provides a link between information theory and statistical physics, by making sure that we add no unconscious arbitrary assumptions about the distribution, even when it satisfies the constraint. We can now introduce the optimization parameters α and $\vec{\theta}$ and solve the following equation as in the paper by Fronczak [1].

$$\frac{\partial}{\partial \mathbb{P}_{\text{can}}(G)} \left[S(\mathbb{P}_{\text{can}}(G)) - \alpha \left(1 - \sum_{G \in \mathcal{G}_n} \mathbb{P}_{\text{can}}(G) \right) \cdots \right.$$
$$\left. - \sum_{i=1}^{n} \theta_i \left(k_i^* - \sum_{G \in \mathcal{G}_n} C_i(G) \mathbb{P}_{\text{can}}(G) \right) \right].$$

This leads to the relation

$$-\log(\mathbb{P}_{\text{can}}(G)) - 1 + \alpha + \sum_{G \in \mathcal{G}_n} \sum_{i=1}^{n} \theta_i C_i(G) = 0$$

and by eliminating the parameter α by including it in the partition function, we obtain P_{can} as defined in Definition 18.

Definition 20 *(Adjacency matrix).* The *adjacency matrix* of a graph G is the $n \times n$ matrix given by

$$g_{ij} = \begin{cases} 0 & \text{if } ij \in E \\ 1 & \text{if } ij \notin E \end{cases}.$$

«

This means that

$$\begin{aligned}
Z(\vec{\theta}) &= \sum_{G\in\mathcal{G}_n} e^{-\vec{\theta}\cdot\vec{C}(G)} \\
&= \sum_{G\in\mathcal{G}_n} \prod_{1\le i<j\le n} e^{-\theta_i g_{ij}(G)} \\
&= \sum_{G\in\mathcal{G}_n} \prod_{1\le i<j\le n} e^{-(\theta_i+\theta_j) g_{ij}(G)} \\
&= \sum_{\substack{G\in\mathcal{G}_n \\ g_{12}=0}} 1 \prod_{\substack{1\le i\le n \\ 2<j\le n}} e^{g_{ij}(G)} + \sum_{\substack{G\in\mathcal{G}_n \\ g_{12}=1}} e^{-(\theta_1+\theta_2)} \prod_{\substack{1\le i\le n \\ 2<j\le n}} e^{g_{ij}(G)} \\
&\vdots \quad \text{continuing this process} \\
&= \prod_{1\le i<j\le n} (1+e^{-(\theta_i+\theta_j)}).
\end{aligned}$$

This allows us to factorize $\mathbb{P}_{\text{can}}$ in the probability of the occurrence of each edge. Namely, put

$$p_{ij} = \frac{e^{-(\theta_i+\theta_j)}}{1+e^{-(\theta_i+\theta_j)}}.$$

Since $1-p_{ij} = \frac{1}{1+e^{-(\theta_i+\theta_j)}}$, it is possible to rewrite $\mathbb{P}_{\text{can}}$ as

$$\mathbb{P}_{\text{can}}(G) = \prod_{1\le i<j\le n} p_{ij}^{g_{ij}} (1-p_{ij})^{1-g_{ij}}.$$

We limit ourselves to the *sparse regime*, where $\max_{1\le i\le n} k_i^* = o(\sqrt{n})$. We compare $\mathbb{P}_{\text{can}}$ to the distribution

$$\hat{\mathbb{P}}_{\text{can}}(G) = \prod_{1\le i<j\le n} \hat{p}_{ij}^{g_{ij}} (1-\hat{p}_{ij})^{1-g_{ij}},$$

where we put $\hat{p}_{ij} = e^{-(\theta_i+\theta_j)} = \frac{k_i k_j}{2L}$ [2]. We first prove measure equivalence of

$\hat{\mathbb{P}}_{\text{can}}(G)$ and $\mathbb{P}_{\text{can}}(G)$. Note that $p_{ij} = \frac{\hat{p}_{ij}}{1+\hat{p}_{ij}}$, to write

$$\begin{aligned}
\frac{1}{n}\log\left(\frac{\hat{\mathbb{P}}_{\text{can}}(G)}{\mathbb{P}_{\text{can}}(G)}\right) &= \frac{1}{n}\sum_{1\le i<j\le n} g_{ij}\log(1+\hat{p}_{ij}) \\
&\quad + \frac{1}{n}\sum_{1\le i<j\le n}(1-g_{ij})\log((1-\hat{p}_{ij})(1+\hat{p}_{ij})) \\
&= \frac{1}{n}\sum_{1\le i<j\le n}\log(1+\hat{p}_{ij}) \\
&\quad + \frac{1}{n}\sum_{1\le i<j\le n}(1-g_{ij})\log(1-\hat{p}_{ij}) \\
&= \frac{1}{n}\sum_{1\le i<j\le n}\log(1+\hat{p}_{ij}) \\
&\quad + \frac{1}{n}\sum_{1\le i<j\le n}\log(1-\hat{p}_{ij}) - \frac{1}{n}\sum_{1\le i<j\le n} g_{ij}\log(1-\hat{p}_{ij}) \\
&= -\frac{1}{n}\sum_{1\le i<j\le n} g_{ij}\log(1-\hat{p}_{ij}) + \frac{1}{n}\sum_{1\le i<j\le n}\log(1-\hat{p}_{ij}^2).
\end{aligned}$$

We want to show that the expression above goes to zero as n goes to infinity. To that end we estimate

$$\begin{aligned}
\left|-\sum_{1\le i<j\le n} g_{ij}\log(1-\hat{p}_{ij})\right| &\le \left|\sum_{1\le i<j\le n} g_{ij}\left(-\hat{p}_{ij}+O(\hat{p}_{ij}^2)\right)\right| \\
&\le \sum_{1\le i<j\le n} g_{ij}\left(\frac{(\max k_i)^2}{2L}+O(\hat{p}_{ij}^2)\right) \\
&\le (\max k_i)^2 + O(\hat{p}^2)
\end{aligned}$$

and

$$0 \le \frac{1}{n}\sum_{1\le i<j\le n}\hat{p}_{ij}^2 = \frac{1}{2}\left[\frac{\sum_{i=0}^n k_i^2}{\sqrt{n}\sum_{i=0}^n k_i}\right]^2 \le \frac{1}{2}\frac{(\max k_i)^2}{n} \to 0.$$

This implies that $\sum_{1\le i<j\le n}\log(1-\hat{p}_{ij}) = -\sum_{1\le i<j\le n} k_i k_j/2L + o(n)$. This is very useful in simplifying the computations. From $\hat{\mathbb{P}}_{\text{can}}$ we obtain

$$\log\hat{P}_{\text{can}}(G) = \sum_{1\le i<j\le n} g_{ij}\log\hat{p}_{ij} + \sum_{1\le i<j\le n}(1-g_{ij})\log(1-\hat{p}_{ij}) \tag{4}$$

$$= \sum_{1\le i<j\le n}\log(1-\hat{p}_{ij}) + \sum_{1\le i<j\le n} g_{ij}\log(\frac{\hat{p}_{ij}}{1-\hat{p}_{ij}}). \tag{5}$$

Substitute $\hat{p}_{ij} = \frac{k_i k_j}{2L}$. We simplify the LHS by working it out as a Taylor

polynomial:

$$\begin{aligned}\sum_{1\le i<j\le n}\log(1-\hat{p}_{ij}) &= -\sum_{1\le i<j\le n}\frac{k_ik_j}{2L}+o(n)\\ &= -\frac{1}{2L}\frac{1}{2}\left(\sum_{i=1}^{n}k_i\sum_{j=1}^{n}k_j-\sum_{i=1}^{n}k_i^2\right)+o(n)\\ &= -\frac{1}{2L}\frac{1}{2}\left(4L^2-\sum_{i=1}^{n}k_i^2\right)+o(n)\\ &= -L+\frac{\overline{k^2}}{2\bar{k}}+o(n).\end{aligned}$$

For the RHS of equation (4) we simplify as follows:

$$\begin{aligned}\sum_{1\le i<j\le n}g_{ij}\log\left(\frac{\hat{p}_{ij}}{1-\hat{p}_{ij}}\right) &= \sum_{1\le i<j\le n}g_{ij}\log\left(\frac{k_ik_j}{2L-k_ik_j}\right)\\ &= \sum_{1\le i<j\le n}g_{ij}\left(\log\left(\frac{k_i}{\sqrt{2L}}\right)+\log\left(\frac{k_j}{\sqrt{2L}}\right)\right)+o(n)\\ &= \sum_{i=1}^{n}k_i\log k_i-L\log(2L)+o(n).\end{aligned}$$

Via Equation (4), noting that $\hat{\mathbb{P}}_{\text{can}}$ and $\mathbb{P}_{\text{can}}$ have the same asymptotic behaviour, we find that

$$\log\mathbb{P}_{\text{can}}(G)=\sum_{i=1}^{n}k_i\log(k_i)-L\log(2L)-L+o(n).$$

For the microcanonical ensemble we calculate the number of configurations of graphs that satisfy the hard constraint. We take an approximation from [5] that gives us the number of graphs $\Omega_{\vec{k}}$ that satisfy the hard constraint

$$\Omega_{\vec{k}}=\frac{\sqrt{2}(\frac{2L}{e})^L}{\prod_{i=1}^{n}k_i!}e^{-(\bar{k}^2/2\bar{k})^2+\frac{1}{4}+o(n^{-1}\bar{k}^3)}.$$

We have $\mathbb{P}_{\text{mic}}(G)=1/\Omega_{\vec{k}}$. The above approximation is only valid when

$\max k_i = o(\sqrt{n})$. Hence

$$\begin{aligned}
\log(\mathbb{P}_{\text{mic}}(G)) &= -\log(\Omega_{\vec{k}}) \\
&= -\log\left(\frac{\sqrt{2}(2L/e)^L}{\prod_{i=1}^n k_i!} e^{-(\overline{k^2}/2\overline{k})^2+\frac{1}{4}+o(n^{-1}\overline{k}^3)}\right) \\
&= \sum_{i=1}^n \log(k_i!) - L\log(\tfrac{2L}{e}) + \left(\frac{\overline{k^2}}{2\overline{k}}\right)^2 + o(n^{-1}\overline{k}^3) \\
&= \sum_{i=1}^n (k_i\log(k_i) - k_i) - L\log(2L) + L + \left(\frac{\overline{k^2}}{2\overline{k}}\right)^2 + o(n^{-1}\overline{k}^3) \\
&= \sum_{i=1}^n k_i\log(k_i) - L\log(2L) - L + \left(\frac{\overline{k^2}}{2\overline{k}}\right)^2 + o(n^{-1}\overline{k}^3)
\end{aligned}$$

It is found that

$$h(\mathbb{P}_{\text{mic}}|\mathbb{P}_{\text{can}}) = \sum_{i=1}^n g(k_i) + o(n).$$

where $g(k) = \log(\frac{k!}{k^k e^{-k}})$ We define the empirical distribution, using the Dirac measure

$$\delta_k(x) = \begin{cases} 1 & x = k \\ 0 & x \neq k \end{cases},$$

as

$$f_n = n^{-1}\sum_{i=1}^n \delta_{k_i^*}.$$

Let f be a degree distribution such that

$$\lim_{n\to\infty} \|f_n - f\|_{\ell^1(g)} = 0,$$

where

$$\|h\|_{\ell^1(g)} = \sum_{k=0}^{\infty} |h(k)|g(k).$$

(Note that $g(k)$ is strictly increasing and $g(0) = 0$.) This means that

$$\lim_{n\to\infty} h(\mathbb{P}_{\text{mic}}|\mathbb{P}_{\text{can}}) = \|f\|_{\ell^1(g)} > 0$$

for any $f \neq \delta_0$. This shows that a graph that has an extensive number of constraints has different asymptotic behavior in the microcanical ensemble than in the canonical ensemble.

4 Conclusion

The Erdõs-Rényi random graph is a very basic model for complex networks. Despite it being very simple, it has a phase transition for the density parameter as shown in the first part. This makes it a good introductory model

for the study of complex networks. The ensemble nonequivalence that was described in the second part shows that the model can be used as an explanation for why ensemble nonequivalence occurs in physical systems. It has this property without the need to introduce so-called *long range interactions*, which were hypothesized in the physics literature to be necessary. This shows that reduction of complex systems to basic examples helps to understand complex systems without making assumptions about their inner workings.

References

[1] Agata Fronczak, *Exponential random graph models* **45** (2007), no. 1994, 2007–2009 pp., available at 1210.7828.

[2] Diego Garlaschelli, Frank den Hollander, and Andrea Roccaverde, *Ensemble nonequivalence in random graphs with modular structure*, Journal of Physics A: Mathematical and Theoretical **50** (2017), no. 1, 015001.

[3] E. T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. **106** (1957May), 620–630.

[4] Claude E Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), no. July 1928, 379–423, available at 9411012.

[5] Tiziano Squartini, Joey de Mol, Frank den Hollander, and Diego Garlaschelli, *Breaking of ensemble equivalence in networks*, Phys. Rev. Lett. **115** (2015Dec), 268701.

[6] Remco Van Der Hofstad, *Random Graphs and Complex Networks, part I (2017)*, Available on http://www.win.tue.nl/rhofstad/NotesRGCN.pdf.